

## IPS Quiz 2

Indicate the answer choice that best completes the statement or answers the question.

	1	2	3	4	5	6	7	8	9	10
a										
b										
c										
d										
e										

1. Plots of the residuals versus fits should show a linear pattern if the regression line is a good fit for your data.

- a. True
- b. False

2. An experiment is conducted to study the bonding strength of adhesives that contain varying amounts of a particular chemical additive. Wafers of a specified material are glued together using the adhesive with each amount of additive and allowed to set for 24 hours, and then the strength needed to separate the wafers is determined. It is reported that the correlation between strength required and amount of additive was 0.86 pound-force per square inch.

Fill in the blanks: This report is \_\_\_\_\_ because correlation must be \_\_\_\_\_.

- a. incorrect; unitless
- b. correct; positive
- c. incorrect; negative
- d. None of the above

3. In order for us to examine the relationship between two variables, the variables must be measured from the same \_\_\_\_\_.

- a. cases
- b. labels
- c. units
- d. values

4. Which of the following statements about a scatterplot is(are) TRUE?

- a. It is always necessary to identify one of the two variables as the explanatory variable and the other as the response variable.

**IPS Quiz 2**

- b. On a scatterplot we look for overall patterns showing the form, direction, and shape of the relationship.
- c. Both A and B are true statements.
- d. None of the above statements is true.

5. Colorectal cancer (CRC) is the third most commonly diagnosed cancer among Americans (with nearly 147,000 new cases annually) and the third leading cause of cancer death (with over 50,000 deaths annually). Research was done to determine whether there is a link between obesity rate and CRC mortality rate among African Americans in the United States by county. Below are the results of a least-squares regression analysis from the software *StatCrunch*.

**Simple linear regression results:**

Dependent Variable: Mortality.rate

Independent Variable: Obesity.rate

Mortality.rate = 13.458199 – 0.21749489 Obesity.rate

Sample size: 3098

R (correlation coefficient) = –0.0067

R-sq = 4.5304943E-5

Estimate of error standard deviation: 111.20661

**Parameter estimates:**

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	13.458199	15.9797735	≠ 0	3096	0.84220207	0.3997
Slope	–0.21749489	0.5807189	≠ 0	3096	–0.37452698	0.708

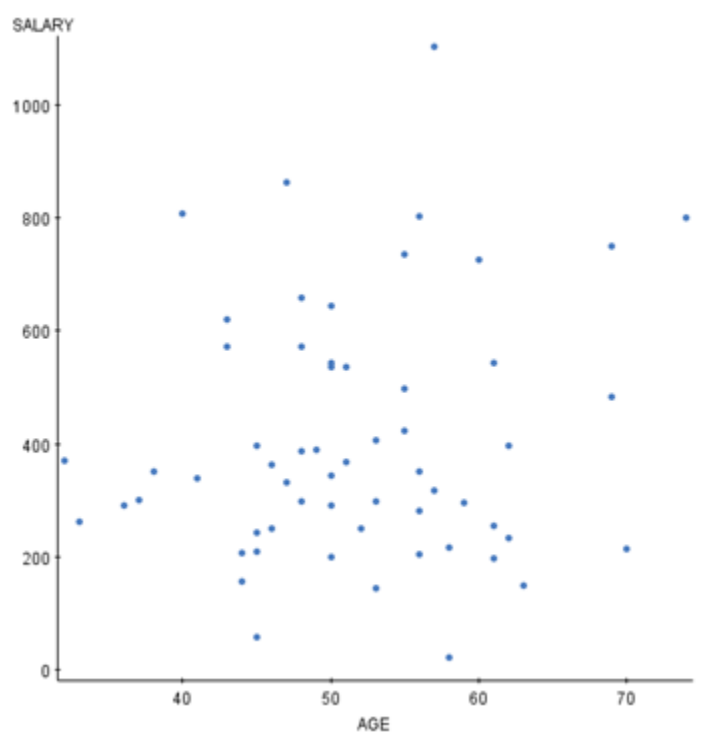
**Analysis of variance table for regression model:**

Source	DF	SS	MS	F-stat	P-value
Model	1	1734.7122	1734.7122	0.14027046	0.708
Error	3096	3.8287952E7	12366.91		
Total	3 9 7	3.8289688E7			

The explanatory variable is \_\_\_\_\_.

- a. obesity rate
- b. CRC mortality rate
- c. slope
- d. intercept

6. Is age a good predictor of salary for a CEO? Sixty CEOs between the ages of 32 and 74 were asked their salary (in thousands). The results of a statistical analysis are shown below.

**IPS Quiz 2****Simple linear regression results:**

Dependent Variable: SALARY

Independent Variable: AGE

 $SALARY = 242.70212 + 3.1327114 \text{ AGE}$ 

Sample size: 59

 $R \text{ (correlation coefficient)} = 0.1276$  $R\text{-sq} = 0.016270384$ 

Estimate of error standard deviation: 220.64246

**Parameter estimates:**

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	242.70212	168.7604	$\neq 0$	57	1.4381461	0.1559
Slope	3.1327114	3.2264276	$\neq 0$	57	0.9709536	0.3357

**Analysis of variance table for regression model:**

Source	DF	SS	MS	F-stat	P-value
Model	1	45896.027	45896.027	0.9427509	0.3357
Error	57	2774936.2	48683.094		
Total	58	2820832.2			

Suppose you wanted to predict the salary of Facebook CEO Mark Zuckerberg, based on the information here. How well do you think your prediction would be, assuming Mr. Zuckerberg was 23 when he started Facebook and became CEO?

- The prediction would be accurate and around \$300,000.
- The prediction would require extrapolation and therefore would not be accurate.

**IPS Quiz 2**

- c. The prediction would be accurate and around \$240,000.  
d. None of the above

7. Colorectal cancer (CRC) is the third most commonly diagnosed cancer among Americans (with nearly 147,000 new cases annually) and the third leading cause of cancer death (with over 50,000 deaths annually). Research was done to determine whether there is a link between obesity rate and CRC mortality rate among African Americans in the United States by county. Below are the results of a least-squares regression analysis from the software *StatCrunch*.

**Simple linear regression results:**

Dependent Variable: Mortality.rate

Independent Variable: Obesity.rate

Mortality.rate = 13.458199 – 0.21749489 Obesity.rate

Sample size: 3098

R (correlation coefficient) = –0.0067

R-sq = 4.5304943E-5

Estimate of error standard deviation: 111.20661

**Parameter estimates:**

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-Value
Intercept	13.458199	15.9797735	≠ 0	3096	0.84220207	0.3997
Slope	–0.21749489	0.5807189	≠ 0	3096	–0.37452698	0.708

**Analysis of variance table for regression model:**

Source	DF	SS	MS	F-stat	P-value
Model	1	1734.7122	1734.7122	0.14027046	0.708
Error	3096	3.8287952E7	12366.91		
Total	3097	3.8289688E7			

What fraction of the variation in mortality rates is explained by the least-squares regression?

- a. 0  
b. 1  
c. –0.0067  
d. 13.45

8. A researcher studies the relationship between “Math SAT score plus Verbal SAT score” and “grade point average (GPA)” for college students at the end of their freshman year. In order to use a relatively homogeneous group of students, the researcher examines only data of high school valedictorians (students who graduated at the top of their high school class) who have completed their first year of college. The researcher finds the correlation between total SAT score and GPA at the end of the freshman year to be very close to 0. Which of the following would be a valid conclusion from these facts?

- a. Because the group of students studied is homogeneous, the results should give an accurate estimate of the correlation the researcher would find if all college students who had completed their freshman year were studied.  
b. If the researcher had studied all college students who had completed their freshman year, the correlation would have been even smaller than that actually found by the researcher. By restricting

## **IPS Quiz 2**

the study to valedictorians, the researcher is examining a group that will be more informative than those students who have completed only their freshman year.

- c. The researcher made a mistake. Correlation cannot be calculated (the formula for correlation is invalid) unless all students who completed their freshman year are included.
- d. None of the above

9. 150. When possible, what is the best way to establish that an observed association is the result of a cause-and-effect relation?

- a. Study the least-squares regression line.
- b. Obtain the correlation coefficient.
- c. Examine  $z$ -scores rather than the original variables.
- d. None of the above

10. Which of the following statements best describes correlation?

- a. Correlation measures whether the relationship between two quantitative variables is linear.
- b. Correlation measures how much of the change in the response variable is caused by a change in the explanatory variable.
- c. Correlation measures the strength of the relationship between any two variables.
- d. Correlation measures the strength of the linear relationship between two quantitative variables.
- e. Correlation measures the strength of the linear association between two categorical variables.

Name: \_\_\_\_\_ Class: \_\_\_\_\_ Date: \_\_\_\_\_

**IPS Quiz 2**

**Answer Key**

1. b
2. a
3. a
4. b
5. a
6. b
7. a
8. d
9. d
10. d